ISCA Archive
http://www.isca-speech.org/archive

TAL 2012 – Third International Symposium on Tonal Aspects of Languages
Nanjing, China, May 26-29, 2012

Domain specificity of lexical tone perception in early infancy: lexical tones versus musical tones

Ao Chen[1], *René Kager*[1]

[1] Institute of Linguistics, Utrecht University, the Netherlands
a.chen@uu.nl, r.w.j.kager@uu.nl

Abstract

In the current study, we tested the perception of pitch variations in early infancy from a domain general perspective for two auditory domains: speech and music. 4-month-old Dutch infants were tested on their discrimination of a musical relative pitch difference, a musical absolute pitch difference, and a Mandarin lexical tone difference. Our results showed that the infants failed to discriminate between Mandarin lexical tones, whereas they succeeded in discriminating both musical pitch differences, both absolute and relative. Crucially, the musical relative pitch difference resembled that between lexical tones, while the acoustical difference between the musical pitches was even smaller than the acoustical difference between lexical tones. Yet, infants only succeeded in discriminating musical pitch contrasts. The performance of the infants who participated in all three tasks conformed to the same pattern. Our findings suggest that young infants are capable of tracking subtle pitch differences, while the lack of proficiency in processing lexical tones is likely to be speech specific and plausibly restricted to lexical tones only.

Index Terms: Mandarin lexical tone, cross-domain perception, perception in early infancy

Introduction

In tone languages, lexical tones are pitch variations used in a phonemic way to distinguish lexical meaning. Acoustically, lexical tones are mainly realized by f0 variations. In Mandarin Chinese, which is the most widely studied tone language, there are four lexical tones, namely the high-level tone (T1), high-rising tone (T2), low-dipping tone (T3), and high-falling tone (T4). A widely accepted way of depicting lexical tones is the tone scale annotation proposed by Chao [1], in which each tone is annotated by numbers from 1 to 5 according to the shape of its f0 contour, where 5 refers to the highest pitch value and 1 to the lowest pitch value. The four lexical tones in Mandarin are described on this tone scale as 55(T1), 35(T2), 214(T3), and 51(T4).

Pitch differences realized by f0 variations are not tone language peculiar. Plentiful f0 information is present in every language by means of intonation or pragmatics. Nevertheless, lexical tones are notoriously difficult for adult L2 learners to acquire. However, infants have been repeatedly observed being able to discriminate between both native and non-native segmental contrasts (e.g. [2] [3]) at birth, and towards the end of the first year of life, their discrimination of non-native phonological contrasts deteriorates while that of native phonological contrasts improves [4]. This perceptual attunement to native language is referred to as perceptual reorganization. Yet, discrepancy still remains regarding the perceptual reorganization of lexical tones. Mattock & Burnham [5][6] found that native English and native French infants were able to discriminate between Thai tones at both 4 months and 6 months, but by 9 months, they were not able to do so anymore. In contrast, native tone language learning infants succeeded in discriminating Thai tones at both 6 and 9 months. Therefore, the authors argued that tonal perceptual reorganization occurs between 6 and 9 months regardless of the prosodic properties (English is stress timed while French is syllable timed) of the infants' native language. Moreover, the same authors [5] showed that the deterioration in discrimination of non-native infants only occurred for lexical tones while the discrimination of the non-speech analogues exhibiting the same pitch contours remained successful from 6 months to 9 months. Liu & Kager [7] tested the discrimination of Mandarin T1 and T4 by native Dutch infants, and they showed that Dutch infants are able to discriminate the well-formed T1-T4 contrast at 5-6 months, 8-9 months, as well as 14-15 months. In another study by Liu & Kager [8], they created an eight-step continuum changing from natural tokens of T1 to T4, and 5-6 months old Dutch infants are able to discriminate the in-between steps which only exhibit less acoustical salient difference in terms of f0. 8-9 months old Dutch infants, however, were only able to discriminate the same in-between steps only if they were trained before the discrimination with a bimodal distribution of the steps that favored the establishment of two contrastive tonal categories, and by 14-15 months, the infants were not able to discriminate the in-between steps even when trained with the bimodal distribution. Chen & Kager [9] tested Dutch infants on their discrimination of Mandarin T2 and T3, and contradictory to the afore-mentioned studies, 6-month-old Dutch infants failed to discriminate this tonal contrast while their discrimination improved to some extent by 9 months. Moreover, unexpectedly, native 4-6 months as well as 11-13 months old Mandarin infants failed to discriminate Mandarin T1-T4 contrast [10]. These confusing results call for a closer examination of early lexical tone perception. Taking into consideration that in the fore-mentioned studies, the stimuli presented to the infants differed substantially in terms of degree of acoustic variation between tokens, and in [5] [6] [8] [9] [10], if the infants were to discriminate between the lexical tones, they had to first normalize the multiple tokens of the same tone presented to them into one single tonal category. As a result, the failure in lexical tone discrimination could be due to the inability to normalize variant tokens rather than the inability to perceive the f0 acoustical differences per se. Therefore, testing infants' perception of lexical tones in terms of pure phonetic differences, i.e. disentangling perception from other cognitive processes such as normalization would help to pinpoint the core obstacle in early lexical tone perception.

Recently, it has been found that young infants are not equally sensitive to all sound contrasts, but they are better at discriminating sounds with salient acoustical difference [11] [12]. With respect to Mandarin lexical tones, it is largely agreed that T2 and T3 are acoustically similar and hence constitute the most difficult contrast perceptually for both native and non-native listeners e.g. [13]. Accordingly, the 6-month-old Dutch infants in [9] failed to discriminate between Mandarin T2 and T3 while 6-month-old Dutch infants did succeed in discriminating Mandarin T1 and T4 could well be due to infants not being able to perceive subtle acoustic differences.

Pitch perception forms an interesting topic for studying human auditory processing in that subtle and precise pitch differences are not only used linguistically in the language domain, such as in lexical tones, but pitch figures prominently in another major auditory domain, music. Two important facets of musical perception involve relative pitch differences and absolute pitch differences respectively, in which the first refers to the ability of detecting melodic contour changes while the latter refers to the ability of detecting the pitch difference when the exactly same melodic contour is shifted to a different pitch height. The enjoyment of music is mainly related to the perception of relative pitch [14]. The intertwined perception of musical and lexical tones has been extensively studied. For example, the rare talent of Absolute Pitch (AP), which refers to the ability to labeling a musical note without presence of other referent note, has been found to occur more frequently among tone language speakers than among non-tone language speakers [15]. English musicians with AP also outperformed non-musicians as well as musicians without AP in discriminating Thai lexical tones [16]. Moreover, it has been shown that musical training and musical aptitude is beneficial for detecting lexical tone differences [17]. In addition, non-tone language listeners with musical perception deficiency performed more poorly on a lexical tone discrimination task compared to normal control groups [18]. Based on these findings, it seems likely that the processing of lexical tones is at least to some extent restricted by a more domain-general limitation on pitch perception.

Infants also possess quite powerful music processing abilities. 6-month-old infants were able to recognize an originally unfamiliar folksong transposed to a different pitch height after being familiarized to that song for a week [19]. In statistical learning tasks in which 8-month-old English infants were trained on continuous musical melodies, they preferred to rely on absolute pitch rather than relative pitch to tell apart tone words and tone part-words according to the transitional probability between the constituent musical notes. However, when absolute pitch was held constant, the infants were able to use relative pitch information to distinguish tone words and tone part-words [20] [21]. Yet, little is known about how very young infants perceive musical pitch differences: do they perceive relative pitch difference more easily like adults or do they resemble older infants in [20] [21] in being able to track absolute pitch differences? Specifically, the absolute pitch difference is presumably more difficult for adult listeners to perceive, and hence the question arises whether young infants are able to discriminate subtle absolute pitch differences.

Bearing in mind that lexical tones are fast and smooth pitch direction changes realized within a short time, while so far no language has been found to have more than five contrastive level tones [22], together with the fact that adults enjoy music mainly by tracking relative pitch changes, it seems evident that relative pitch difference is perceptually more salient naturally. If so, we would expect infants to discriminate musical relative pitch more easily than musical absolute pitch, and if relative pitch difference is perceptually salient, then we would expect infants to discriminate lexical tones that differ in pitch contour too. Moreover, most Western music is composed in semitones, with a minimum pitch interval of one semitone, which equals the pitch difference between adjacent piano keys. In comparison, in the language domain, pitch values of lexical tones such as the onset and offset of the tonal contours, always differ by more than one semitone both within a lexical tone and across different lexical tones. Therefore, if lexical tone discrimination by young infants is restricted by a general auditory limitation in perceiving subtle pitch changes, and if acoustical saliency is a crucial factor to ensure successful discrimination, then if infants fail to discriminate lexical tones, we would not expect them to discriminate the more subtle musical pitch differences.

To explore the issue of lexical tone perception in early infancy from a domain general perspective, in the current study, we tested native Dutch 4-month-old infants on their discrimination of an acoustically not so salient linguistic contrast, namely Mandarin lexical tones T2 and T3, as well as musical pitch differences, relative and absolute. Importantly, in order to examine the role of acoustical saliency in lexical tone discrimination, infants in the tone discrimination experiment were tested with one single token of each tonal category rather than multiple tokens to reduce the cognitive demand of normalization. A subgroup of the participants participated in all three experiments, and their data are analyzed separately.

2. Experiments

2.1 Experiment 1 Mandarin lexical tone discrimination

2.1.1 Participants

21 4-month-old healthy Dutch infants (age range 4: 02-4:29) participated in this experiment. None of the infant had been diagnosed with hearing deficiency or had reported ear infection. Among the 21 participants, 11 were girls and 10 were boys. Another infant was tested and excluded for analysis for crying.

2.1.2 Stimuli

To generate the stimuli, first a native female Mandarin speaker recorded the syllable /ma/ carrying T2 and T3 respectively together with other syllables carrying different lexical tones. Among the multiple productions of the /ma/ T2 and /ma/ T3, one token of each tone was selected for further manipulation. In order to make sure that the two syllables only differ in lexical tone while other acoustical properties such as intensity and duration were equal, the two syllables were manipulated as follows in Praat [23]. First the f0 contour of each syllable was extracted respectively, and 100 points at equal temporal distance were selected along each extracted contour. Then we replaced the f0 values of the 100 points along the T3 contour with that of the 100 points of T2 at original time points, and regenerated a pitch contour of T2 according to the newly established 100 points. Next we replaced the original T3 f0 contour with the newly generated T2 contour, and finally we re-synthesized the original /ma/T3 with the newly generated T2 f0 contour, thus obtaining a new /ma/T2. The original /ma/T3 and the new /ma/T2 were used as stimuli in the experiment. By manipulating the naturally produced tokens in this way, we assure that the manipulated /ma/T2 and /ma/T3 are identical in segmental spectrogram, intensity, and duration. Both of the tones have a duration of 445ms. Multiple native speakers of Mandarin listened to the manipulated stimuli, and no one reported perceptual ambiguity or unnaturalness. The f0 contour of the two tones is shown in Figure 1. As could be read from the figure, the lowest points along the f0 of T2 and T3 differ by at least 5 semitones, the offsets of the f0 of the two tones differ by at least 5 semitones, and the maximum pitch differences between the two contours is at least 6 semitones.

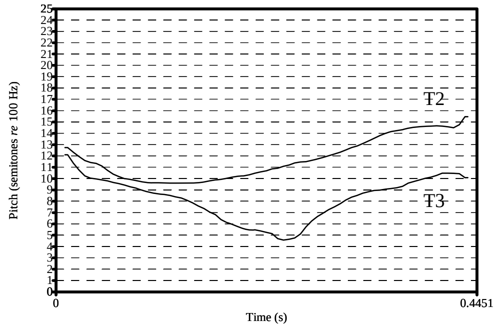

Figure 1. f0 contours of the stimuli /ma/T2 and /ma/T3 in semitones.

2.1.3 Procedure

The visual fixation paradigm was selected for the current study. A test cabin and a separate control room for the experimenter were used. During the experiment, infants sat on their parent's lap in the test cabin, in front of a 14 inch computer screen displaying the visual stimuli about one meter away from the baby. The auditory stimuli were presented at a comfortable volume through a hidden speaker in front of the baby. The parent listened to background music through headphones to prevent possible interaction with the infants. There was a hidden camera above the screen recording the looking behavior of the infants, and the video was transferred to experimenter's computer in the control room. The experimenter observed and recorded the visual fixation of the infants by pressing the "looking" and "non-looking" button on a button box connected to the control computer.

The habituation phase was followed by the test phase, and a pretest and a posttest were used to measure general attention of the participants, in which the stimuli were moving infant friendly pictures accompanied by beeps. When the pretest finished, and once the participant focused on the screen, the experimenter initiated the first habituation trial by pressing the "looking" button, and once the participant looked way, the experimenter pressed the "non-looking" button. The looking and non-looking of the participants were always recorded by these two buttons. If the infant looked away and looked back to the screen within two seconds, the same trial continued, and if the infant looked away for more than two seconds, the trial ended and a smiling baby face appeared on the screen in order to regain the participant's attention. Once the infant looked back to the screen, the experimenter started the next trial by pressing the "looking" button again. The looking time of each look and total looking time per trial were recorded automatically by the experimenter's control computer.

The total looking time over the first three trials in the habituation phase was used as a baseline to measure habituation. Once the total looking time of three consecutive trials dropped below 65% of the total looking time of the first three habituation trials, the habituation criterion was met, and the test phase started automatically. In the test phase, infants were presented with one "old" trial with /ma/ bearing the same tone as they had heard in the habituation phase, and a "novel" trial with /ma/ carrying the tone that they had not heard yet in the habituation phase. The tones that were used in habituation phase and the order of the "old" and "novel" trials in the test phase were counter-balanced among the participants.

If the infants were able to detect the difference between the two tones, then in the test phase, upon hearing the novel trial, their listening time should recover due to hearing something new, i.e. they would have a longer looking time to the visual stimuli when presented with the novel trial than when presented with the old trial.

2.1.4 Results and discussion

The videos of the participants were recoded offline after the experiment before submitting the data to analysis. After recoding, the raw looking times of the "old" and "novel" trials were logarithmically converted to correct the skewness of the distribution of the raw looking time data. The log transformed looking times fitted a normal distribution ($D_{LGLT\ old}$ (21) = 0.123, p>0.1; $D_{LGLT\ novel}$ (21) = 0.091, p>0.1), and the statistics hereafter are based on the log transformed looking time (LGLT).

A 1-tailed paired T test was carried out between LGLT of the "old" and "novel" trials. The pair did not reveal a significant difference: $T_{old\text{-}novel}$ (20) = –0.334, p>0.05. Figure 2 gives the average LGLT of novel stimuli and the average LGLT of old stimuli. As can be seen in the figure, compared to the looking time when presented with the old trials, the infants did not look longer to the visual stimuli when they were presented with new stimuli, suggesting that they were not able to discriminate between Mandarin T2 and T3.

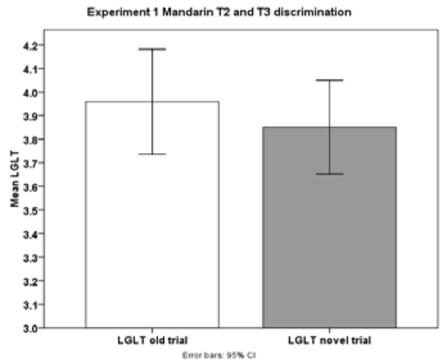

Figure 2 Mean LGLT of the old trial and the novel trial of Experiment 1.

2.2 Experiment 2 Musical relative pitch discrimination

2.2.1 Participants

22 healthy 4 months old (4:02-4:29) participated in the current experiment. None of the infant has been diagnosed with hearing deficiency or was reported to have ear infection. Among the 22 participants, 11 are girls and 11 are boys. Another four infants were tested and excluded for analysis for crying.

2.2.2 Stimuli

16th notes of D4, E4, F4, and C4 were synthesized using Nyquist script (for a description of Nyquist, see http://audacity.sourceforge.net/help/nyquist and http://www.cs.cmu.edu/~music/music.software.html). The notes were generated with a timbre of piano, and all the notes were generated on the C4 (middle C) scale, along which the fundamental frequency of A4 equals to 440Hz. Nyquist synthesized the notes with the default equal temperament, i.e. the pitch differences between adjacent notes, such as between D4 and #D4, between E4 and F4 are equal to one semitone. After synthesizing the four single notes separately, D4, E4, F4 were concatenated to get a three-note melody D4E4F4 (melody 1), and D4, C4, and F4 were concatenated to get another three-note melody D4C4F4 (melody 2). Melody 1 and melody 2 were used as stimuli in this experiment. Comparing melodies 1 and 2, we can see that in terms of pitch contours, melody 1 has a rising direction while melody 2 has a dipping pitch contour, or in other words, relative pitch patterns differ between melodies 1 and 2. Importantly, melodies 1 and 2 were controlled to have identical initial as well as final pitch levels, so that if infants were to discriminate between the melodies, they would not succeed by only paying attention to the initial or final portion of the melody. Rather, they had to perceive each melody of three notes as a whole and discriminate them based on the complete pitch contour. Melodies 1 and 2 both have a duration of 830ms. Pitch contours of melodies 1 and 2 are given in Figure 3.

2.2.3 Procedure

Exactly the same procedure as in Experiment 1 was used for this experiment. The only difference is that the auditory stimuli were changed to melody 1 and melody 2. In the habituation phase, infants were habituated on one of the two melodies, and in the test phase, they were tested with two trials, in which in the old trial they were presented with the same melody as they had heard in habituation, and in the novel trial, they were presented with the other melody that they had not hear yet. The habituation melody and the order of old trial and the novel trial in the test phase were counterbalanced among the infants.

If infants were able to discriminate between these two melodies, then in the test phase, their looking time to the visual stimuli was supposed to be recovered due hearing something new.

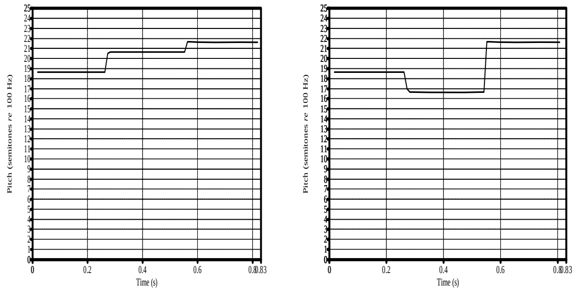

Figure 3 Pitch contours of melody 1 (upper panel) and melody 2 (lower panel) in semitones.

2.2.4 Results

The videos of all participants were recoded offline after the experiment before submitting the data to analysis. After recoding, the raw looking times of the “old” and “novel” trials were logarithmically converted to correct the skewness of the distribution of the raw looking time data. The log transformed looking times fitted a normal distribution ($D_{LGLT\ old}$ (22) = 0.094, p>0.1; $D_{LGLT\ novel}$ (22) = 0.156, p>0.1), and the statistics hereafter are based on the log transformed looking time (LGLT).

A 1-tailed paired T test was carried out between LGLT of the “old” and “novel” trials. The pair reveals a significant difference: $T_{old\text{-}novel}$ (21) = −1.942, $p<0.05$. Figure 4 gives the average LGLT of novel stimuli and the average LGLT of old stimuli. As can be seen from the figure, the mean LGLT of the novel trial is significantly longer than the mean LGLT of the old trial. It serves as evidence that listening time of the infants increased when they heard a new melody in the test phase, and hence the infants were able to discriminate the two melodies that differ in relative pitch.

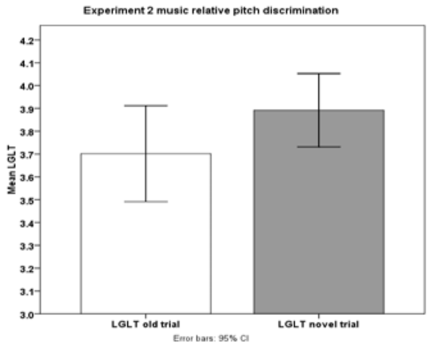

Figure 4 *Mean LGLT of the old trial and the novel trial in Experiment 2.*

2.3 Experiment 3 Musical absolute pitch discrimination

The acoustical difference between the two melodies in Exp. 2 were five semitones. In order to find out whether 4-month-old infants are able to discriminate between melodies with even more subtle acoustical difference, Experiment 3 tested infants on two melodies that only differ in one semitone.

2.3.1 Participants

17 healthy 4-month-old (4:02-4:29) native Dutch infants participated in the experiment. None of the infants had been diagnosed with hearing deficiency or was reported to have ear infection. Among the 17 participants, 8 are girls and 9 are boys.

2.3.2 Stimuli

Another pair of 16^{th} notes #D4 and #F4 were generated in Nyquist with the same default. Next, #D4, F4, and #F4 were concatenated to obtain another three-note melody #D4F4#F4 (melody 3). In this case, the pitch contours of both melodies have rising directions. Importantly, the pitch intervals between each note are exactly the same for both melodies, i.e. 2 semitones between the first and the second note, and 1 semitone between the second note and the third note. In other words, the two melodies have identical relative pitch shapes but differ in one semitone in terms of absolute pitch height. Both the melodies have a duration of 830ms. The f0 contours of the melodies are given in Figure 5.

2.3.3 Procedure

Exactly the same procedure as in Experiment 1 was used for this experiment. The only difference was that the auditory stimuli were changed to melody 1 and melody 3. In the habituation phase, infants were habituated on one of the two melodies, and in the test phase, they were tested with two trials, in which the in the old trial they were presented with the same melody as they had heard during habituation, and in the novel trial, they were presented with the other melody that they had not heard yet. The habituation melody and the order of old trial and the novel trial in the test phase were counterbalanced among the infants.

If infants were able to discriminate between these two melodies, then in the test phase, their looking time to the visual stimuli should recover due to hearing something new.

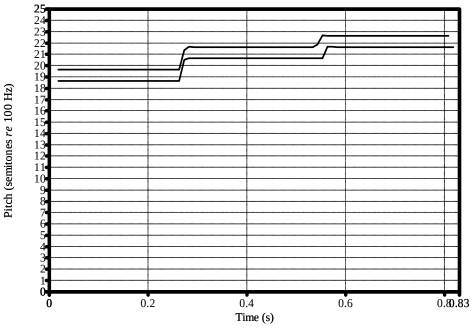

Figure 5 Pitch contours of melody 1 (lower line) and melody 3 (upper line) in semitones.

2.3.4 Results and discussion

The videos of all participants were recoded offline after the experiment before submitting the data to analysis. After recoding, the raw looking times of the "old" and "novel" trials were logarithmically converted to correct the skewness of the distribution of the raw looking time data. The log transformed looking times fitted a normal distribution ($D_{LGLT\ old}$ (17) = 0.106, p>0.1; $D_{LGLT\ novel}$ (17) = 0.153, p>0.1), and the statistics reported hereafter are based on the log transformed looking time (LGLT).

A 1-tailed paired T test was carried out between LGLT of the "old" and "novel" trials. The pair reveals a significant difference: $T_{old\text{-}novel}$ (16) = –2.122, p<0.05. Figure 6 gives the average LGLT of novel stimuli and the average LGLT of old stimuli. As can be seen from the figure, the mean LGLT of the novel trial is significantly longer than the mean LGLT of the old trial. It serves as evidence that the listening time of the infants increased when they heard a new melody in the test phase, and hence the infants were able to discriminate the two melodies that differ in absolute pitch.

2.4 Cross-experiment analysis

Among the infants that participated in Experiments 1, 2 and 3, there were 16 that participated in all three experiments (8 boys and 8 girls). Each of these 16 infants was tested on two separate days. They always participated in one music experiment and the lexical tone discrimination experiment on one day, and then participated in the other music experiment combined with another speech experiment on the other day. The order of music experiment and lexical tone experiment for each infant was fixed across test dates, i.e. if on experiment day one the infant first participated in music experiment and then followed the lexical tone experiment, then on experiment day 2 he or she also first participated in the music experiment and then followed another speech experiment. For one single infant, there were at least three days between the two test dates. The order of the experiments across test dates was counterbalanced for each infant. However, due to the limited number of participants, within each test date, 6 of them first participated in a music task and 10 of them first participated in the lexical tone task.

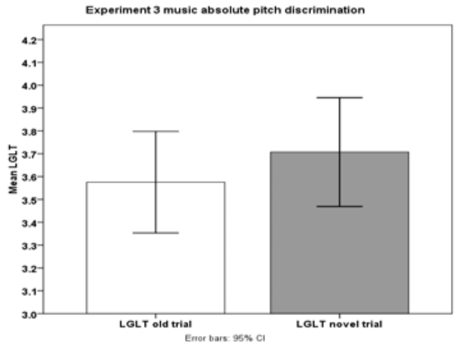

Figure 6 *Mean LGLT of the old trial and the novel trial in Experiment 2.*

Among these 16 infants, the same patterns were observed for Experiment 1, Experiment 2, and Experiment 3. A 1-tailed T test showed that the LGLT of the novel trial is significantly longer in Experiment 2 ($T_{old\text{-}novel}$ (15) = -1.820, p<0.05) and in Experiment 3 ($T_{old\text{-}novel}$ (15) = -1.781, p<0.05), while it failed to show significance difference in LGLT in Experiment 1 ($T_{old\text{-}novel}$ (15) = 0.902, p>0.1). The mean LGLT of these 16 participants in Experiment 1, Experiment 2, and Experiment 3 is given in Figure 7. Moreover, a repeated measures ANOVA revealed significant main effect for experiment (Experiment 1, 2, and 3) ($F_{experiment}$ (2, 15) = 4.217, p < 0.05), but not for trial types (old, novel) ($F_{trial\ types}$ (1, 15) = 1.445, p>0.1,), and not the interaction between experiment and trial type ($F_{experiment*trial\ types}$ (2, 30) = 2.452, p>0.1). Admittedly, there are only 16 participants, so to test more infants would probably amplify the interaction effect.

The cross-experiment analysis has largely eliminated the possible influence from individual difference, and the result of the repeated measures ANOVA serves as convincing evidence that the 4-month-old infants behaved differently on the music tasks and the lexical tone task. Taken together with what can be seen in Figure 7, it is clear that 4-month-old Dutch infants discriminated musical melodies more easily than Mandarin lexical tones. The fact that the general looking time to the lexical tones is longer than that to musical melodies could be due to that the majority of the infants first participated in the lexical tone task, and therefore their general attention is higher than in the following music task.

3. General discussion

In the current study, we tested Dutch 4-month-old infants on their discrimination of Mandarin lexical tones, musical relative pitch, and musical absolute pitch. It was found that the infants successfully discriminated musical pitch differences of both a relative and absolute nature. However, infants failed to show discrimination of Mandarin lexical tone T2 and T3.

Regarding the acoustical properties of musical stimuli and lexical tone stimuli, first, the duration of the musical stimuli was much longer than that of the lexical tone stimuli, while in the musical melodies, there were three independent elements. As a result, presumably, the memory load needed for processing the musical stimuli would be larger than that needed for processing the lexical tone stimuli. However, even when dealing with a cognitively less demanding task, i.e. lexical tone discrimination, infants still performed poorly as compared to the musical pitch discrimination task. This pattern suggests that as young as 4 months, infants are already perceiving pitch patterns realized in music and pitch patterns realized in speech differently.

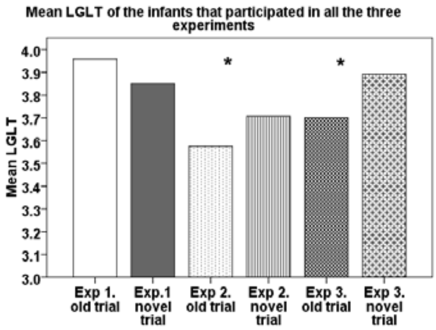

Figure 7 Cross-experiment Mean LGLT of old and novel trial of the infants that participated in all three experiments.

A crucial finding of the current study is that infants as young as 4 months are already perceptually well equipped to perceive subtle pitch differences: they are able to track as subtle as one semitone pitch difference. However, this fine-grained perception of pitch is domain specific. Even though Mandarin T2 and T3 differ for at least 5 semitones in terms of f0 lowest point, f0 offset, and maximum f0 difference, the infants still failed to discriminate between them. Bearing in mind that in Experiment 1, the infants were tested with one single token of Mandarin T2 and one single token of Mandarin T3, which is purely phonetic discrimination as in Experiment 2 and 3**,** it can be concluded that the failure in Experiment 1, as well as discrimination difficulties observed in previous lexical tone discrimination studies [10] [11] cannot be simply attributed to a general auditory failure in perceiving acoustically not-so-salient pitch differences. It is more likely that the perception of lexical tones is restricted by some speech-specific limitations. In [9], the 5-6 month old infants succeeded in discriminating Mandarin T1 and T4, and the f0 offsets of the stimuli representing these tones differ more than 10 semitones. Hence possibly, at least for infants, it seems that the minimal pitch difference needed for successful discrimination of lexical tones has to be larger than the minimum difference need for music pitch discrimination.

Another point that is worth noting is that, in Experiment 2, melodies 1 and 2 were generated in such a way that their pitch shapes resembled the f0 contours of Mandarin T2 and T3. Admittedly, the degree of pitch rising and falling is different in music melodies and Mandarin T2/T3, nevertheless, melody 1 and T2 (tone scale 35) both have a rising contour, and melody 3 and T3 (tone scale 214) both have a dipping contour. Hence, the failure in Experiment 1 cannot be due to infants not being able to track the difference between a rising and dipping pitch contour, but again it seems that the failure in lexical tone discrimination is speech specific and even possibly lexical tone specific.

Undoubtedly, pitch patterns realized in music differ much from f0 contour in lexical tones, such that pitch change across different notes in music is abrupt but the f0 contours of lexical tones are continuous and smooth, and that within a musical note the pitch is stable while in lexical tones the f0 contour varies continuously, and it would be informative for future research to find out how, if any, cross-domain perception of pitch is restricted by universal auditory biases.

4. References

[1] Chao, Y.,-R. “Fanqie ye ba zhong (eight variaties of languages based on the principle of fanqie).” Bulletin of the Institute of History and Philology, Academia Sinica, *2*, 320-354, 1931

[2] Werker, J. F., and Tees, R. C. “Cross-language speech perception: Evidence for perceptual reorganization during the first year of life.” Infant Behavior and Development, *7*(1), 49-63, 1984

[3] Polka, L., and Werker, J. F. “Developmental changes in perception of nonnative vowel contrasts.” Journal of Experimental Psychology, *20*(2), 421-435, 1994

[4] Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., and Iverson, P. “Infants show a facilitation effect for native language phonetic perception between 6 and 12 months.” Developmental Science, *9*(2), F13-F21, 2006

[5] Mattock, K., & Burnham, D. “Chinese and English infants' tone perception: Evidence for perceptual reorganization.” Infancy, 10(3), 241-265, 2006

[6] Mattock, K., Molnar, M., Polka, L., & Burnham, D. “The developmental course of lexical tone perception in the first year of life.” Cognition, 106(3), 1367-1381, 2008

[7] Liu, L. and Kager, R., “Is perceptual reorganization affected by statistical learning? Dutch infants' sensitivity to lexical tones”, BUCLD35 Proceedings, Vol.2, 404-413, Cascadilla Press, 2011.

[8] Liu, L. and Kager, R., “Non-native tone perception from infant to adult: How consistent and flexible is it?” Speech Prosody Proceedings, 2012

[9] Chen, A., and Kager, R. “Lexical tone perception in early infancy.” Speech Prosody, 6th International Conference, 2012

[10] Shi, R., Gao, J., and Li, A. “Perception of lexical tones in infants.” 36th BUCLD proceeding, 2011

[11] Chandan R., N. “The acoustic–perceptual salience of nasal place contrasts.” Journal of Phonetics, 36(1), 191-217, 2008

[12] Narayan, C. R., Werker, J. F., and Beddor, P. S. “The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination.” Developmental Science, 13(3), 407-420, 2010

[13] Hume, E., and Johnson, K. “The impact of partial phonological contrast on speech Perception.”, Manuscript, 2003

[14] Trehub, S. E., and Hannon, E. E. “Infant music perception: Domain-general or domain-specific mechanisms?” Cognition, 100(1), 73-99, 2006

[15] Deutsch, D., Henthorn, T., and Dolson, M. “Absolute pitch, speech and tone language: Some experiments and a proposed framework.” Music Perception, 21(3), 339-356. 2004

[16] Burnham, D., and Brooker, R. “Absolute pitch and lexical tones: Tone perception by non-musician, musician, and absolute pitch non-tonal language speakers.” Proceedings of 7th International Conference on Spoken Langauge Processing, Denver, Colorado. 257-260, 2002

[17] Wang, P., & Perrachione, T. “Learning pitch patterns in lexical identification by native English-speaking adults.” Applied Psycholinguistics, 28, 565-585, 2007

[18] Nan, Y., Sun, Y., and Peretz, I. “Congenital amusia in speakers of a tone language: Association with lexical tone agnosia.” Brain, 133(9), 2010

[19] Plantinga, J., and Trainor, L. J. “Memory for melody: Infants use a relative pitch code.” Cognition, 98(1), 1-11, 2005

[20] Saffran, J. R., and Griepentrog, G. “Absolute pitch in infant auditory learning: Evidence for developmental reorganization.” Developmental Psychology, 37(1), 74-85, 2001

[21] Saffran, J. R., Johnson, E. K., Aslin, R. N., and Newport, E. L. “Statistical learning of tone sequences by human infants and adults.” Cognition, 70(1), 27-52, 1999

[22] Blevins, J. “Evolutionary phonology: The emergence of sound patterns.” Cambridge: Cambridge University Press, 2004

[23] Boersma, P., & Weenink, D. Praat: Doing phonetics by computer (version 5.1.05)